

**PERAMALAN KUALITAS UDARA DI KOTA MAKASSAR
MENGUNAKAN METODE AUTOREGRESSIVE INTEGRATED
MOVING AVERAGE (ARIMA) DAN AUTOREGRESSIVE
INTEGRATED MOVING AVERAGE WITH EXOGENOUS
VARIABLES (ARIMAX)**

Disusun untuk memenuhi Tugas Mata Kuliah: Statistika



Dosen Pengampu:
Farhanna Mar'i, S.Kom., M.Kom.

Penyusun :

Dimas Dwi Pramono Nugroho	23051204329
M. Galih Rahmatulloh	23051204343
Pebri Andika Putra	23051204334
Muh. Farhan Baidlowi	23051204352

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS NEGERI SURABAYA
TAHUN AKADEMIK 2025**

DAFTAR ISI

BAB I PENDAHULUAN.....	3
1.1 Latar belakang	3
1.2 Rumusan masalah.....	3
1.3 Tujuan penelitian	4
1.4 Manfaat.....	4
BAB II LANDASAN TEORI	5
2.1. Kualitas udara	5
2.2 Uji Augmented Dickey-Fuller (ADF).....	5
2.3 Uji Auto Correlation Function & Partial Auto Correlation Function (ACF & PACF).....	6
2.4 Metode ARIMA (Autoregressive Integrated Moving Average)	7
2.5 Metode ARIMAX (Autoregressive Integrated Moving Average)	10
2.6 Evaluasi Model Peramalan	10
BAB III METODOLOGI.....	11
3.1 Perolehan dan Pengolahan Data	11
3.2 Tahapan analisis	15
3.3 Alat bantu	18
BAB IV HASIL DAN PEMBAHASAN	19
4.1 Analisis Eksplorasi	19
4.2 Model ARIMA.....	19
4.3 Pembangunan model ARIMAX	25
4.4 Evaluasi model ARIMAX dan ARIMA.....	31
BAB V KESIMPULAN	32
5.1 KESIMPULAN.....	32
5.2 SARAN.....	32
DAFTAR PUSTAKA	33

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kualitas udara merupakan salah satu indikator penting untuk menilai kesehatan lingkungan suatu wilayah (Moshammer & Wallner, 2011). Saat ini, Kota Makassar mengalami peningkatan yang cukup signifikan dalam aktivitas industri, transportasi, dan pertumbuhan penduduk. Hal tersebut mengakibatkan peningkatan emisi polutan di udara (Ramdhan & Syamsuddin, 2020). Salah satu polutan utamanya ialah PM 2.5, yaitu partikel halus dengan diameter kurang dari 2,5 mikrometer. Partikel ini cukup berbahaya bagi kesehatan dikarenakan ukuran yang sangat kecil sehingga dapat menembus sistem pernapasan hingga mencapai alveolus paru-paru. Hal ini dapat mengakibatkan gangguan kesehatan yang cukup serius, seperti gangguan pernapasan dan penyakit kardiovaskular (Keramidas et al., 2023).

Fluktuasi kualitas udara di Kota Makassar dapat dilihat berdasarkan data indeks Standar Pencemar Udara (ISPU) di beberapa titik, seperti di Terminal Daya dan Terminal Malengkeri yang tercatat dalam kategori tidak sehat (101-200) bahkan berbahaya (>300) (Jusri, 2023; Utama, 2019). Selebihnya, studi prediktif dengan data pada tahun 2020 hingga 2022 memperkirakan bahwa pada tahun 2045 rata-rata ISPU di Kota Makassar akan meningkat hingga 105. Hal ini menunjukkan bahwa kualitas udara memerlukan perhatian serius dalam pengelolaan dan pengendalian polusi udara (Jusri, 2023).

Peningkatan aktivitas manusia yang tidak diimbangi dengan pengelolaan lingkungan efektif dapat menyebabkan penurunan kualitas udara yang berdampak negatif pada kesehatan masyarakat dan lingkungan. Oleh karena itu, strategi dalam mengelola kualitas udara yang komprehensif dan data prediktif diperlukan untukantisipasi dan mengurangi dampak buruk dari polusi udara di masa mendatang.

1.2 Rumusan Masalah

1. Bagaimana prediksi pola harian PM2.5 di Makassar menggunakan model ARIMA?
2. Bagaimana pengaruh faktor meteorologi terhadap prediksi harian PM2.5 di Makassar menggunakan model ARIMAX?

1.3 Tujuan Penelitian

1. Merancang dan menganalisis prediksi pola harian PM2.5 di Makassar menggunakan model ARIMA
2. Merancang dan menganalisis pengaruh faktor meteorologi terhadap prediksi harian PM2.5 di Makassar menggunakan model ARIMAX.

1.4 Manfaat Penelitian

1. Memberikan gambaran awal tentang bagaimana metode statistik dapat digunakan dalam memprediksi fenomena lingkungan.
2. Menjadi referensi awal bagi mahasiswa atau peneliti lain yang ingin mengembangkan model prediksi kualitas udara yang lebih kompleks.

BAB II

LANDASAN TEORI

2.1 Kualitas Udara

Kualitas udara menggambarkan tingkat kebersihan udara berdasarkan kandungan polutan di dalamnya. Beberapa parameter utama yang sering digunakan untuk mengukur kualitas udara antara lain: PM2.5, PM10, SO₂, NO₂, dan CO (WHO, 2021). PM2.5 adalah partikel mikroskopis yang berukuran lebih kecil dari 2.5 mikrometer, yang dapat masuk ke saluran pernapasan manusia sehingga menyebabkan gangguan kesehatan (Pope & Dockery, 2006; Chen et al., 2020).

2.2 Uji Augmented Dickey-Fuller (ADF)

Uji Augmented Dickey-Fuller (ADF) adalah metode umum yang sering diterapkan dalam uji stasioneritas data deret waktu (time series) untuk deteksi keberadaan root pada data (Basuki & Prawoto, 2016; Widarjono, 2018). Unit root mengindikasikan bahwa data bersifat non-stasioner dengan sifat statistik seperti rerata dan varians yang berubah tiap waktu. Sehingga, perlu penerapan transformasi sebelum analisis lebih lanjut.

Uji ADF melakukan perhitungan mengenai autokorelasi pada residual dengan menambahkan lag variabel dependen sebagai variabel independen dalam model regresi. Tujuannya ialah untuk menghilangkan autokorelasi serial pada error term sehingga hasil uji lebih valid (Repository UIN Suska, 2022). Berikut adalah perhitungan secara matematisnya:

$$\Delta Y_t = \alpha_1 + \alpha_2 T + \gamma Y_{t-1} + \sum_{i=1}^m \beta_i \Delta Y_{t-1} + \varepsilon_t$$

Dengan :

- Δ : adalah operator differencing pertama,
- Y_t : adalah nilai variabel pada waktu t,
- T : adalah tren waktu,
- m : adalah jumlah lag diterapkan dalam menghilangkan autokorelasi,
- ε_t : adalah error team.

Uji dilakukan dengan uji hipotesis nol $H_0: \gamma = 0$, yaitu data mengandung unit root atau non-stasioner dibandingkan dengan $H_a: \gamma < 0$, yaitu data stasioner. Keputusan dalam pengujian berdasar pada perbandingan statistik ADF terhadap nolai kritis dari distribusi MacKinnon. Apabila nilai absolut statistik ADF lebih besar dari nilai kritis, maka hipotesis nol ditolak dan data merupakan stasioner (Akbar, 2016; Widarjono, 2018). Uji ADF juga dapat diterapkan dengan memanipulasi intercept dan tren deterministik tergantung karakteristik data yang diuji (Digilib Unila, 2017). Apabila data tidak stasioner, maka dilakukan defferencing hingga stasioner. Hal ini

merupakan bagian dari proses permodelan deret waktu, seperti ARIMA (Repository ITK, 2020). Oleh karena itu, uji ADF merupakan langkah awal yang krusial dalam pemodelan data time-series.

2.3 Uji Koefisien Korelasi

Uji koefisien korelasi merupakan salah satu metode statistik yang digunakan untuk melihat sejauh mana hubungan antara dua variabel. Melalui uji ini, kita dapat mengetahui apakah dua variabel memiliki hubungan yang positif, negatif, atau justru tidak memiliki hubungan sama sekali. Kemudian Salah satu bentuk koefisien korelasi yang paling umum digunakan adalah koefisien korelasi Pearson. Uji ini digunakan ketika data bersifat interval atau rasio dan berdistribusi normal. Nilai koefisien korelasi (r) berada pada rentang -1 hingga +1, yang masing-masing memiliki makna sebagai berikut:

- Nilai r mendekati +1 menunjukkan hubungan positif yang kuat, artinya jika satu variabel naik maka variabel lainnya juga cenderung naik.
- Nilai r mendekati -1 menunjukkan hubungan negatif yang kuat, artinya jika satu variabel naik maka variabel lainnya cenderung menurun.
- Nilai r mendekati 0 menandakan bahwa hubungan antara kedua variabel tersebut lemah atau tidak signifikan.

Menurut Sugiyono (2015), interpretasi nilai koefisien korelasi secara umum dapat dijelaskan sebagai berikut:

Tabel 1. Korelasi

Nilai r	Keterangan
0,00 – 0,199	Hubungan sangat lemah
0,20 – 0,399	Hubungan lemah
0,40 – 0,599	Hubungan cukup
0,60 – 0,799	Hubungan kuat
0,80 – 1,000	Hubungan sangat kuat

Uji ini sering digunakan dalam penelitian kuantitatif untuk melihat seberapa besar pengaruh atau keterkaitan antara dua variabel. Selain itu, uji korelasi juga kerap digunakan untuk menguji validitas instrumen, khususnya pada tahap uji coba angket atau kuesioner, guna memastikan bahwa setiap item pertanyaan benar-benar mengukur aspek yang dimaksud.

2.4 Uji Auto Correlation Function & Partial Auto Correlation Function (ACF & PACF)

Autocorrelation Function (ACF) merupakan alat statistik yang diterapkan untuk

melakukan pengukuran kekuatan hubungan antara nilai data pada waktu ke-t dengan nilai data pada periode sebelumnya (lag). ACF diterapkan dalam melakukan identifikasi model Autoregressive Integrated Moving Average (ARMA) dengan memerhatikan korelasi antar pengamatan pada waktu berbeda. Hal ini sangat penting untuk menentukan apakah data memiliki pola tertentu atau sudah bersifat acak (white-noise). Berikut adalah perhitungan matematis pada lag ke-K:

$$\rho_k = \frac{\sum_{t=1}^{n-k} (Z_t - Z_{rata}) (Z_{t+k} - Z)}{\sum_{t=1}^{n-k} (Z_t - Z)^2}$$

Dengan:

- Z_t : data pada waktu ke-t
- Z_{rata} : rerata pada data
- n : jumlah data

Partial Autocorrelation Function (PACF) merupakan alat statistik yang diperuntukkan untuk melakukan pengukuran korelasi antara nilai data pada waktu ke-t dengan nilai data pada waktu ke-($t + k$) melalui eliminasi pengaruh nilai-nilai antara keduanya (lag 1 hingga lag k-1). PACF diterapkan dalam mengetahui hubungan langsung diantara dua titik waktu setelah menghilangkan pengaruh perantara agar sehingga dapat membantu menentukan orde model AR (Autoregressive). Berikut adalah persamaan Yule-Walker untuk PACF pada lag ke-K:

$$\phi_{kk} = \text{corr}(Y_t, Y_{t+k} | Y_{t+1}, \dots, Y_{t+k-1})$$

Atau dalam bentuk matriks, seperti berikut:

$$\Gamma_{k\phi k} = \rho_k$$

ACF diterapkan untuk mengecek pola musiman, tren, dan keacakan pada data. Plot ACF yang menurun secara perlahan mengindikasikan bahwa data belum stasioner dan perlu differencing. PACF diterapkan untuk menentukan orde model AR. Berikut adalah interpretasinya:

- Apabila ACF menurun perlahan dan PACF terpotong pada lag tertentu, maka data kemungkinan cocok untuk model AR.
- Apabila ACF terpotong pada lag tertentu dan PACF menurun secara perlahan, maka data kemungkinan cocok untuk model MA.

2.5 Metode ARIMA (Autoregressive Integrated Moving Average)

Metode Autoregressive Integrated Moving Average (ARIMA) adalah teknik dalam analisis deret waktu yang umum diterapkan dalam peramalan data runtun waktu

menggunakan pola tren dengan musiman (Box & Jenkins, 1976). Secara struktur, ARIMA memadukan tiga komponen, yaitu autoregressive (AR), integrated (I), dan Moving Average (MA) sehingga memungkinkan pemodelan tidak stasioner melalui proses differencing (Peramalan pencemaran udara di Pekanbaru, 2015). Model ARIMA biasanya dinotasikan sebagai ARIMA (p, d, q) . Berikut adalah rincian penjelasannya:

- p merupakan orde komponen autoregressive.
- d merupakan orde differencing, yaitu banyaknya iterasi dalam differencing agar data stasioner.
- q merupakan orde dari komponen moving average.

Berikut adalah perhitungannya secara matematis (Wahyuni & Setyawan, 2020):

$$\phi_p B \phi_p B^s (1 - B)^d (1 - B^s)^D Y_t = \theta_q(B) \theta_q(B^s) a_t$$

Dengan:

- B : operator *backshift*
- $\phi_p B$: koefisien komponen AR non musiman dengan derajat P ,
- $\phi_p B^s$: koefisien komponen AR musiman S ,
- $(1 - B)^d$: operator untuk *differencing* orde d ,
- $(1 - B^s)^D$: operator untuk *differencing* musiman S orde D ,
- $\theta_q(B)$: koefisien komponen MA non musiman dengan derajat q .
- $\theta_q(B^s)$: koefisien komponen MA musiman S ,
- a_t : nilai residu pada waktu ke- t .

Model ARIMA diidentifikasi dengan melihat plot *autocorrelation function* (ACF) dan plot *partial autocorrelation function* (PACF) yang telah dilakukan uji stasioner. Orde proses ARIMA ditentukan dengan lag yang keluar pada plot ACF, dan plot PACF. Berikut adalah tabel identifikasi model ARIMA:

Tabel 2. ACF dan PACF

Model	ACF	PACF
AR (p)	Dies Down (Turun cepat secara eksponensial/sinusoidal)	Cuts off setelah lag q
MA (q)	Cuts off setelah lag q	Dies Down (Turun cepat secara eksponensial/sinusoidal)
ARMA (p,q)	Turun cepat setelah lag (q-p)	Turun cepat setelah lag (p-q)
AR (p) atau MA (q)	Cuts off setelah lag q	Cuts off setelah lag q
White noise (Random)	Tidak ada yang signifikan (tidak ada yang keluar batas)	Tidak ada yang signifikan (tidak ada yang keluar batas)

2.6 Metode ARIMAX (Autoregressive Integrated Moving Average)

Metode ARIMAX adalah pengembangan dari ARIMA yang memasukkan variabel-variabel eksogen, yaitu variabel luar yang dianggap memiliki pengaruh terhadap variabel utama yang diprediksi (Jurnal JTIC, 2024). Dalam konteks ini, kualitas udara dapat dipengaruhi oleh variabel eksogen seperti suhu dan kelembaban (Hyndman, 2022). Kemudian berikut ini merupakan model matematisnya:

$$y_t = \delta t + \sum_g^4 \mathbf{1}_{y_g} V_{g,t} + \sum_{i=1}^{12} \beta_i M_{i,t} + \frac{\theta q(B) \theta Q(B^s)}{\phi_p B \phi_p B^s (1-B) d(1-B^s) \dot{D}} a_t$$

2.7 Evaluasi Model Peramalan

Evaluasi model peramalan adalah tahap penting dalam mengukur seberapa baik hasil prediksi yang diperoleh melalui model yang mendekati nilai aktual. Dengan melakukan evaluasi, nilai akurasi dan keandalan model dalam prediksi data di masa depan. Beberapa metrik evaluasi yang umum digunakan dalam analisis peramalan antara lain *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*

2.7.1 MSE (Mean Absolute Error)

MSE (Mean Absolute Error) merupakan rumus yang digunakan untuk menghitung rata-rata dari kuadrat selisih antara nilai aktual y_i dan nilai prediksi \hat{y}_i . Kemudian Karena menggunakan kuadrat, MSE memperbesar kesalahan besar dan sensitif terhadap outlier, jadi jika nilai MSE lebih kecil maka menandakan model lebih akurat. Berikut ini merupakan rumus dari MSE :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

2.7.2 RMSE (Root Mean Square Error)

RMSE adalah akar kuadrat dari MSE, sehingga satuannya sama dengan satuan data asli, kemudian RMSE juga menekankan kesalahan besar karena berasal dari kuadrat selisih. Selain itu rumus ini diterapkan dalam melihat seberapa jauh prediksi dari nilai sebenarnya dalam skala yang sama. Berikut ini merupakan rumus dari model

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

2.7.3 MAPE (Mean Absolute Percentage Error)

MAPE merupakan rumus yang digunakan untuk mengukur rata-rata kesalahan absolut dalam bentuk persentase terhadap nilai aktual, yang dimana cocok untuk membandingkan performa model pada data berskala besar. Namun rumus ini tidak cocok jika data mengandung nilai 0 (ini dikarenakan akan menyebabkan pembagian nol). Berikut ini merupakan rumus dari MAPE :

$$MAPE = \sum_{t=1}^n \left| \frac{(y_i - \hat{y}_i)}{\hat{y}_i} \right| \times 100\%$$

BAB III METODOLOGI PENELITIAN

3.1 Flowchart (Alur Penelitian)

Berikut adalah tahapan alur penelitian proyek kami:

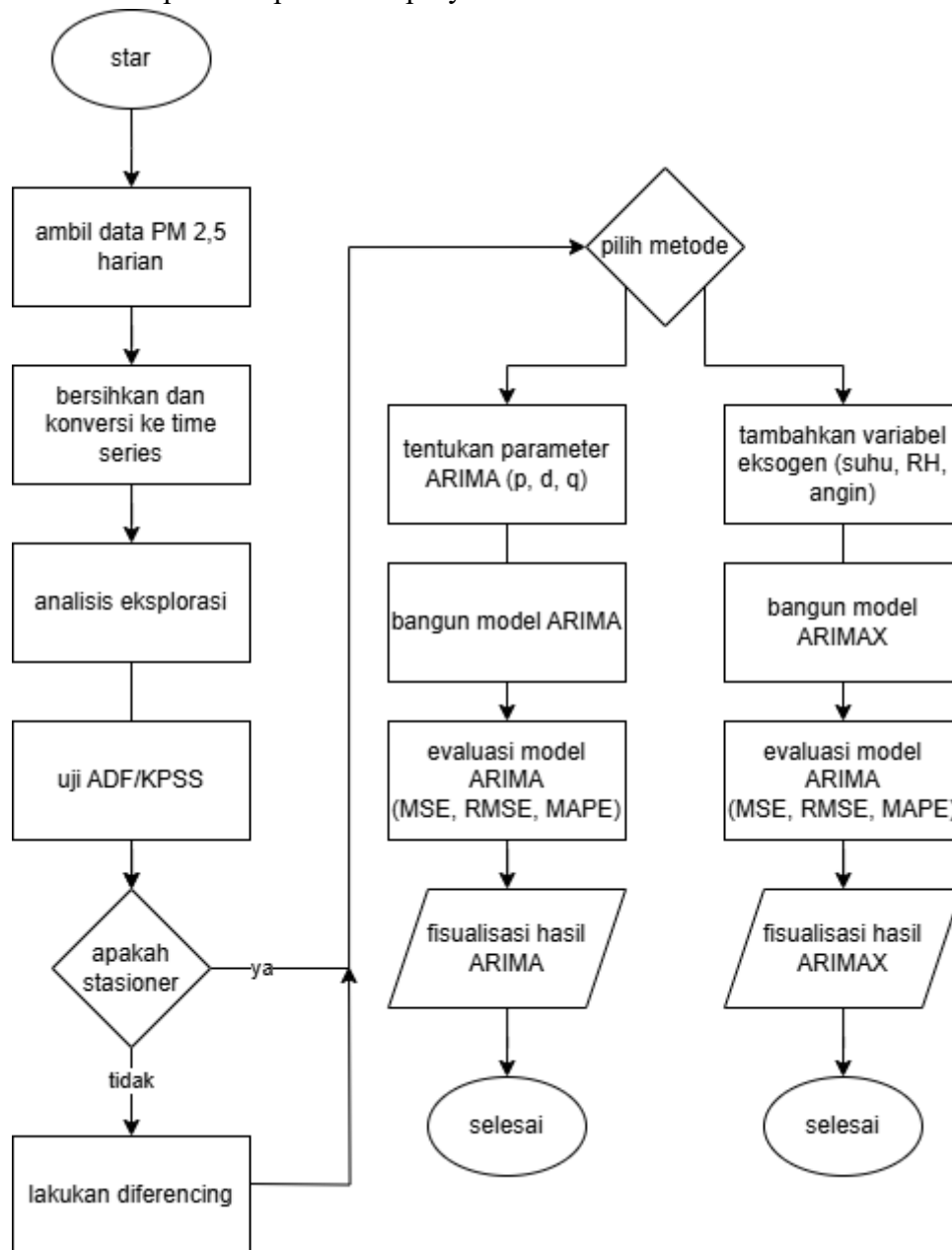


Fig 1. Alur Penelitian

3.2 Perolehan dan Pengolahan Data

Penelitian ini menggunakan dua jenis data utama, yaitu data kualitas udara (PM 2.5) & data iklim (suhu udara, kelembapan, & kecepatan angin) untuk di kota Makassar.

Pemilihan Kota Makassar sebagai lokasi penelitian didasarkan pada tren peningkatan kualitas udara yang memburuk dalam lima tahun terakhir, seperti yang tercatat dalam laporan pemantauan kualitas udara regional. Sebagai salah satu kota metropolitan terbesar di Indonesia timur, Makassar mengalami pertumbuhan pesat sektor transportasi dan pembangunan, yang berkontribusi terhadap peningkatan emisi partikel halus di udara. Berikut adalah deskripsi sumber dan pengolahan data:

3.2.1 Data PM 2.5 Makassar

Data yang digunakan dalam penelitian ini terdiri dari data harian konsentrasi PM2.5 dan data parameter meteorologi di Kota Makassar. Data PM2.5 diperoleh dari platform AQICN (Air Quality Index China) melalui API resmi mereka (<https://aqicn.org>), sedangkan data meteorologi seperti suhu udara, kelembapan, dan kecepatan angin diperoleh dari BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) melalui situs web resmi.

Data konsentrasi harian PM2.5 diperoleh melalui situs Air Quality Index Project (AQICN) melalui API publik. Pengambilan data diterapkan secara otomatis dengan pemrograman API untuk menarik informasi harian dari stasiun pemantauan udara di Makassar. Kemudian, data yang telah diperoleh disimpan dalam format tabular untuk keperluan analisis lebih lanjut.

Tabel 3. Data PM2.5 Makassar

TANGGAL	PM2.5						
	Min	Max	Median	Q1	Q3	Standar Deviasi	Count
Friday, 7 February 2025	5.5	6.81	6.04	5.72	6.51	0.417	21
Saturday, 8 February 2025	5.25	7.83	6.13	5.69	7.07	0.765	24
Sunday, 9 February 2025	7.58	8.21	7.89	7.81	7.97	0.146	24
Monday, 10 February 2025	7.58	8.75	8.07	7.82	8.5	0.38	24
Tuesday, 11 February 2025	7.93	10.91	9.44	8.76	10.58	1.006	23
Wednesday, 12 February 2025	9.78	11.56	10.19	10.07	10.76	0.528	24
Thursday, 13 February 2025	8.19	10.81	9.82	8.91	10.57	0.893	24
Friday, 14 February 2025	8	9.94	8.52	8.44	9.22	0.539	24
Saturday, 15 February 2025	8.56	10.06	9.03	8.77	9.31	0.435	24
Sunday, 16 February 2025	8.73	12.44	9.61	9.38	10.43	1.076	24
Monday, 17 February 2025	11.46	12.94	12.65	11.95	12.8	0.468	24
Tuesday, 18 February 2025	9.06	11.6	10.37	10.18	10.96	0.664	24
Wednesday, 19 February 2025	6.06	9.03	8.9	6.81	8.94	1.175	24
Thursday, 20 February 2025	4.23	6.1	5.4	5.18	5.52	0.437	24

Friday, 21 February 2025	4.34	4.75	4.61	4.45	4.71	0.151	4
Monday, 24 February 2025	5.23	6.48	5.96	5.59	6.13	0.353	12
Tuesday, 25 February 2025	4.69	6.27	5.16	4.98	5.28	0.425	24
Wednesday, 26 February 2025	4	6.62	6.34	5.39	6.44	0.853	24
Thursday, 27 February 2025	3.79	7.67	4.38	4.23	4.65	0.965	18
Friday, 28 February 2025	8.13	8.56	8.19	8.16	8.28	0.146	6
Saturday, 1 March 2025	5.46	7.62	6.38	5.5	7.56	0.911	17
Sunday, 2 March 2025	7.67	11.45	10.02	8.68	10.31	1.116	24
Monday, 3 March 2025	10.47	12.62	11.91	11.68	12.33	0.556	24

3.2.2 Periode Data

Rentang waktu yang dianalisis adalah Februari hingga Mei 2025, yang secara strategis dipilih karena mencerminkan peralihan dari musim hujan ke musim kemarau di wilayah Makassar. Periode ini sangat penting dalam studi kualitas udara karena pada masa transisi musim, fluktuasi cuaca cenderung mempengaruhi konsentrasi partikel polutan seperti PM_{2.5}, terutama karena peningkatan kelembapan, berkurangnya curah hujan, dan meningkatnya aktivitas transportasi masyarakat setelah musim hujan berakhir.

Periode observasi data dimulai dari tanggal 7 Februari 2025 hingga 24 Mei 2025 mencakup lebih dari tiga bulan data harian. Rentang waktu dipilih untuk melakukan pengamatan hubungan diantara variabel iklim dan konsentrasi PM_{2.5} serta melakukan regresi.

3.2.3 Data Iklim Harian BMKG Makassar

Data iklim harian yang diterapkan meliputi:

- Suhu rerata harian (TAVG)
- Kelembapan rerata harian (RH_AVG)
- Suhu Minimum (TN)
- Suhu Maksimum (TX)
- Curah Hujan (RR)
- Sinar Matahari (SS)
- Kecepatan Angin Maks (FF_X)
- Kecepatan Angin Rata (FF_AVG)
- Arah Angin Dominan (DDD_CAR)

Data yang digunakan diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Makassar dalam bentuk Excel yang berisi konten hasil pengamatan meteorologis harian. Berikut ini merupakan contoh data yang digunakan:

Tabel 4. Data Iklim BMKG Makassar

TANGGAL	TN	TX	TAVG	RH_AVG	RR	SS	FF_X	DDD_X	FF_AVG	DDD_CAR
01/02/2025	24	31	26.6	87	27.8	2	6	30	2	E
02/02/2025	25	31	27.8	82	9.6	5	5	320	2	N
03/02/2025	26	32	28.7	73	0	8	3	290	1	NW
04/02/2025	26	31	28	80	0	5	5	330	1	NW
05/02/2025	26	31	28	80	0.6	3	5	300	2	NW
06/02/2025	25	30	26.5	87	10.4	5	6	300	2	SE
07/02/2025	24	27	25.1	92	33.6	0	7	310	3	NE
08/02/2025	24	29	26	91	77.2	0	8	280	5	W
09/02/2025	24	29	24	96	44	4	8	280	3	E
10/02/2025	23	30	27.2	89	89	0	9	290	6	W
11/02/2025	25	29	25.7	94	28	2	8	280	3	E
12/02/2025	24	30	26	91	99.2	0	8	310	2	E
13/02/2025	24	31	27.2	87	7	1	5	310	2	E
14/02/2025	25	31	27.7	85	2.2	8	3	310	2	E
15/02/2025	25	31	27.6	84	1.8	2	7	270	2	E
16/02/2025	25	30	27.4	83	0	7	3	310	1	E
17/02/2025	25	30	27.2	86	0	2	3	310	1	E

3.2.4 Penanganan Data Kosong

Dalam proses pengolahan data, ditemukan adanya nilai yang hilang (missing values) baik pada data PM2.5 maupun variabel meteorologi. Kehadiran missing value merupakan kendala serius dalam analisis deret waktu karena dapat mengganggu proses pemodelan dan menghasilkan estimasi yang bias jika tidak ditangani dengan tepat.

Untuk mengatasi hal ini, dilakukan imputasi data menggunakan metode Hybrid Mean Interpolation, yaitu kombinasi interpolasi linier sederhana dengan rata-rata lokal untuk memperkirakan nilai yang hilang berdasarkan observasi sebelum dan sesudahnya. Metode ini dipilih karena mempertahankan kesederhanaan model namun juga cukup efektif dalam menangkap pola musiman atau fluktuasi halus dalam data deret waktu.

Alasan Pemilihan Hybrid Mean Interpolation:

1. Dibandingkan interpolasi linier murni, metode ini mampu menghindari estimasi yang terlalu tajam pada data yang fluktuatif.
2. Lebih efisien secara komputasi dibandingkan dengan linear interpolasi.
3. Tidak memerlukan pelatihan model tambahan seperti metode regresi atau time series forecasting untuk estimasi data hilang.

Namun, metode ini memiliki kelemahan dalam menangani missing data yang terjadi berturut-turut dalam jumlah besar, karena interpolasi bisa menghasilkan nilai yang tidak mencerminkan fluktuasi alami data. Oleh karena itu, metode ini hanya

diterapkan pada data yang memiliki missing value dalam jumlah terbatas dan tidak berurutan.

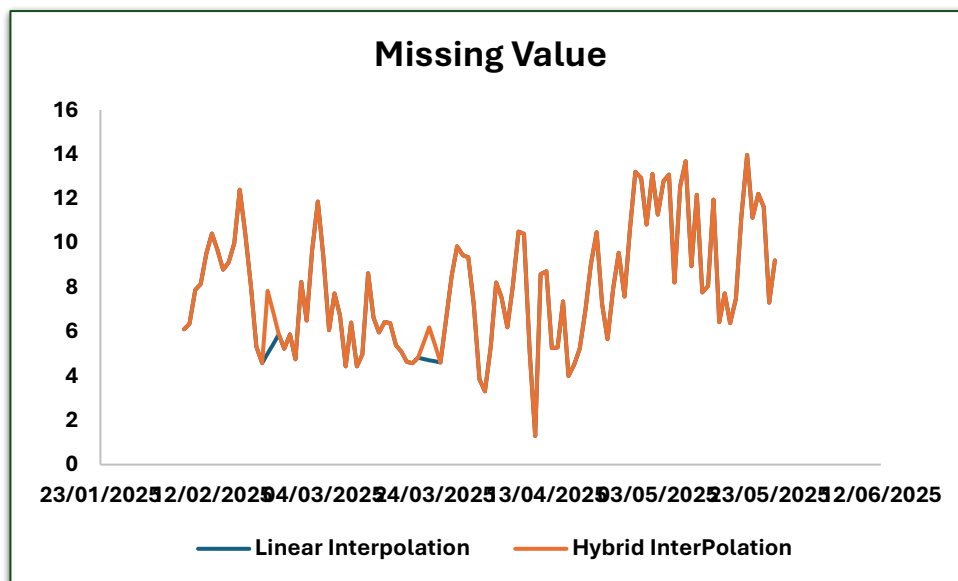


Fig 2. Missing Value Linear Interpolasi dan Hybrid Interpolasi

Berdasarkan hasil eksplorasi awal, total data yang hilang pada rentang waktu Februari–Mei 2025 adalah sekitar 2,9% dari total observasi, dengan distribusi yang tersebar secara acak dan tidak membentuk pola missing block yang signifikan. Hal ini mendukung penerapan metode imputasi berbasis interpolasi lokal.

Dalam proses pengumpulan data, didapat beberapa nilai yang hilang atau kosong (missing values). Demi menjaga kontinuitas data deret waktu dan mengisi kekosongan tersebut, diterapkan metode Hybrid Mean Interpolation yaitu gabungan dari:

Tabel 5. Perhitungan Missing Value PM2.5

PM2.5
6.10625
6.3575
7.89125
8.13875
9.55
10.4225
9.7
8.7875
9.105
10.00125
12.4

10.46
8.04875
5.31625
4.57875
(Data Ke-17 Missing)
(Missing)
5.88375

Menentukan nilai kedua berurutan yang hilang dengan menggunakan mean:

Data Ke 17	7.8416875
	6.86271875

Menentukan nilai pertama berurutan yang hilang dengan menggunakan interpolasi:

Dengan rumus

$$Data\ Ke - 17 = \frac{PM_{16} + PM_{18}}{2}$$

Sehingga didapat hasil sebagai berikut:

5.31625
4.57875
7.8416875
7.8416875
5.88375

3.3 Tahap Analisis

Secara umum, pada proses analisis ini, terdapat beberapa tahap diantaranya sebagai berikut:

3.3.1 Analisis Eksplorasi

Pada tahap pertama, dilakukan dengan proses pengunduhan data kualitas udara harian (khususnya PM2.5) dan data cuaca (seperti suhu, kelembaban, dan kecepatan angin) dari sumber terpercaya. Setelah data diperoleh, dilakukan proses penggabungan data berdasarkan tanggal agar membentuk satu kerangka data yang utuh. Kemudian selanjutnya dilakukan eksplorasi data untuk memahami pola umum, outlier, serta tren

musiman atau harian yang muncul. Setelah dilakukan eksplorasi data untuk memahami pola, maka visualisasi seperti plot garis dan boxplot digunakan untuk mengamati distribusi nilai dan fluktuasi kualitas udara harian di Makassar.

3.3.2 Permodelan ARIMA

Model *AutoRegressive Integrated Moving Average* (ARIMA) diterapkan untuk melakukan pemodelan dan peramalan data *time-series* PM 2.5. Berikut adalah tahapan dalam membangun model ARIMA:

1. Uji Stasioneritas
Melakukan uji *Augmented Dickey-Fuller* (ADF) untuk tahu apabila data bersifat stasioner. Apabila data tidak stasioner, maka *differencing* diterapkan agar supaya data menjadi stasioner.
2. Identifikasi Model
Pola *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) dilakukan analisis untuk menentukan parameter p (*autoregressive*), d (*differencing*), dan q (*moving average*).
3. Estimasi dan Evaluasi Model
Beberapa kombinasi model dianalisis, mulai dari ARIMA (0,1,1) hingga ARIMA (2,1,2) dan dilakukan pemilihan model terbaik berdasarkan nilai-nilai evaluasi, seperti RMSE (*Root Mean Squared Error*), MAPE (*Mean Absolute Percentage Error*), N-BIC (*Normalized Bayesian Information Criterion*), dan hasil uji *white-noise* menggunakan Ljung-Box Q-Test.
4. Pemodelan dan Peramalan
Model ARIMA terbaik diterapkan dalam melakukan peramalan nilai PM2.5 selama 16 hari ke depan.

3.3.3 Permodelan ARIMAX

Dalam meningkatkan tingkat akurasi peramalan, dilakukan pengembangan model ARIMA dengan penambahan variabel eksogen, yaitu Tavg: temperatur rerata (°C), RH_avg: kelembapan rerata (%), ff_avg: kecepatan angin rerata (m/s) sehingga model ARIMA *with Exogenous Variables* (ARIMAX). Berikut adalah tahapannya:

1. Persiapan Data Eksogen
Data suhu, kelembapan, dan angin diselaraskan secara temporal dengan data PM2.5. Proses ini mencakup penggabungan data berdasarkan tanggal dan normalisasi jika diperlukan, agar skala antarvariabel tidak terlalu berbeda dan model tetap stabil.
2. Persiapan Data Eksogen
Model dibangun menggunakan fungsi `Arima()` dari R, dengan memasukkan variabel eksogen sebagai argumen tambahan. Setelah model terbentuk, dilakukan evaluasi menggunakan MSE, RMSE, dan MAPE untuk menilai performa prediksi dan dibandingkan dengan model ARIMA murni.

3. Evaluasi Model ARIMAX
4. Peramalan ARIMAX

3.3.4 Visualisasi dan Interpretasi Hasil

Hasil akhir dari proses analisis divisualisasikan dalam bentuk grafik deret waktu, yang menampilkan data aktual dan hasil prediksi dari model. Visualisasi ini membantu melihat seberapa dekat hasil prediksi mengikuti pola data asli. Selain itu, dilakukan interpretasi terhadap koefisien model dan pengaruh variabel eksogen, untuk memberikan pemahaman yang lebih mendalam terhadap dinamika kualitas udara di Makassar.

3.4 Alat Bantu

Pada pemrosesan penganalisan data, adapun bantuan tools yang digunakan seperti :

1. R studio
2. Microsoft Excel
3. SPSS

BAB IV HASIL DAN PEMBAHASAN

4.1 Analisis Eksplorasi

Untuk memastikan kualitas hasil imputasi, dilakukan visualisasi data sebelum dan sesudah imputasi. Pada gambar dibawah ini dihasilkan korelasi antar variabel dalam data:

		PM25	Temperatur Minimum	Temperatur Maksimum	Temperatur Rata-Rata	Kelembaban Relatif Rata-Rata	Curah hujan	Sinar Matahari	Kecepatan Angin	Arah Angin	Kecepatan Angin Rata-Rata
PM25	Pearson Correlation	1	.359**	.294**	.412**	-.357**	-.023	.142	-.175	-.112	-.122
	Sig. (2-tailed)		.000	.003	.000	.000	.819	.155	.079	.263	.224
	N	102	102	102	102	102	102	102	102	102	102

Fig 3. Korelasi Antar Variabel

Data diperoleh dari file Excel yang berisi informasi harian mengenai konsentrasi PM2.5 dan variabel meteorologi: suhu rata-rata (TAVG), kelembapan relatif rata-rata (RH_AVG), dan kecepatan angin rata-rata (FF_AVG). Data dibersihkan dengan menghapus entri tidak lengkap dan dikonversi ke format deret waktu untuk analisis lanjutan.

Tabel 6. Data yang diambil dalam penelitian

TANGGAL	TAVG	RH_AVG	FF_AVG	PM2.5
07/02/2025	25.1	92	3	6.10625
08/02/2025	26	91	5	6.3575
09/02/2025	24	96	3	7.89125
10/02/2025	27.2	89	6	8.13875
11/02/2025	25.7	94	3	9.55
12/02/2025	26	91	2	10.4225
13/02/2025	27.2	87	2	9.7
14/02/2025	27.7	85	2	8.7875
15/02/2025	27.6	84	2	9.105
16/02/2025	27.4	83	1	10.0013
17/02/2025	27.2	86	1	12.4
18/02/2025	27.4	85	2	10.46
19/02/2025	27.4	87	2	8.04875
20/02/2025	26.7	89	2	5.31625
21/02/2025	26.9	85	2	4.57875

4.2 Implementasi ARIMA

4.2.1 Model ARIMA

Pembuatan model ARIMA dirancang dengan dua perangkat lunak, yaitu R Studio & SPSS. Pendekatan ganda ini diterapkan agar mempercepat pengerjaan, meminimalkan kesalahan, dan melakukan penyederhanaan proses analisis. R Studio

diterapkan untuk eksplorasi data dan melakukan pemodelan fleksibel. Selain itu, SPSS diterapkan untuk melakukan validasi statistik dan melakukan pengujian performa model secara numerik. Berikut adalah flowchart dari program:

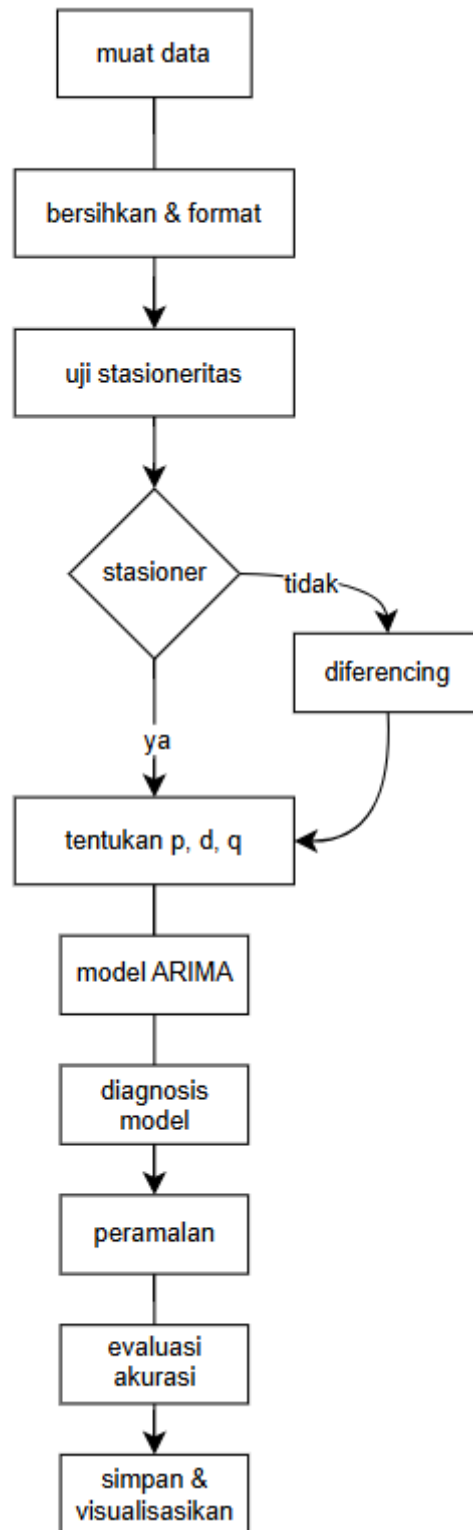


Fig 4. Model ARIMA

4.2.2 Implementasi ARIMA

Setelah data PM2.5 dilakukan konversi menjadi time series, pengujian stasioneritas menggunakan Augmented Dickey-Fuller (ADF Test). Hasil yang diperoleh menunjukkan bahwa data belum stasioner, maka dilakukan differencing sebanyak satu kali. Setelah differencing pertama data dinyatakan telah stasioner.

Selanjutnya identifikasi model dilakukan melalui analisis ACF (Autocorrelation Function) dan PACF (Partial Autocorrelation Function). Kemudian model diuji secara iteratif mulai dari ARIMA (0,1,0) hingga ARIMA (2,1,2).

Tabel 7. Model Arima

Model ARIMA	Error			Variabel Signifikan	Variabel Non-Signifikan	Ujung Box Q-Test	
	MAPE	RMSE	N-BIC			P Value	White-Noise
(0,1,1)	27.709	2.243	1.703	1	0	0.049	White-Noise
(0,1,2)	25.004	2.078	1.595	1	0	0.542	Tidak
(1,1,0)	27.397	2.308	1.761	0	1	0.002	White-Noise
(1,1,1)	26.76	2.124	1.638	2	1	0.114	Tidak
(1,1,2)	25.028	2.087	1.647	1	2	0.545	Tidak
(2,1,0)	26.165	2.167	1.679	0	2	0.297	Tidak
(2,1,1)	26.418	2.17	1.725	1	2	0.285	Tidak
(2,1,2)	26.427	2.118	1.721	2	2	0.117	Tidak

Terpilih

Berdasarkan hasil yang diperoleh, model ARIMA (0,1,2) dipilih karena memberikan performa yang paling baik.

4.2.3 Perhitungan Model ARIMA (0,1,2)

Model ARIMA (0,1,2) merupakan model yang tidak memiliki komponen *autoregressive* (AR), memiliki satu kali *differencing* (I) dan dua komponen *moving average* (MA). Notasi umum model ARIMA (0,1,2) ialah:

$$(1 - B)y_t = (1 + \theta_1 B + \theta_1 B^2)\varepsilon_t$$

Dengan:

B = operator backshift

y_t = data aktual

ε_t = residual (white noise)

$\theta_1 \theta_2$ = parameter untuk MA orde 1 dan orde 2. Berdasarkan hasil analisis di SPSS:

θ_1 = 0.290

θ_2 = 0.501

Sehingga model matematisnya menjadi, seperti berikut:

$$y_t - y_{t-1} = \varepsilon_t + 0.290\varepsilon_{t-1} + 0.501\varepsilon_{t-2}$$

Berikut adalah interpretasi dari model:

- Perubahan nilai PM2.5 pada waktu ke-t tergantung pada gangguan acak (*error*) saat ini dan dua periode sebelumnya.
- Karena komponen AR tidak ada, maka nilai data tidak secara langsung diterapkan untuk prediksi. Namun, hanya perubahan nilainya yang diodelkan melalui MA(1) dan MA (2).
- Koefisien MA (1) dan MA (2) signifikan berdasarkan *p-value*, yaitu 0,001 dan 00,0 yang menunjukkan bahwa kedua komponen ini memberikan kontribusi nyata dalam membentuk model yang prediktif.

Berikut adalah simulasi prediksinya:

Misalkan nilai perubahan PM2.5 sebelumnya:

- $\varepsilon_{t-1} = 0.8$
- $\varepsilon_{t-2} = -0.5$
- $\varepsilon_t = 0.6$

Maka:

$$\begin{aligned}\Delta y_t &= 0.6 + (0.290 \times 0.8) + (0.501 \times -0.5) \\ &= 0.6 + 0.232 - 0.2505\end{aligned}$$

$$\Delta y_t = 0.5$$

Jika $y_{t-1} = 34$, maka:

$$y_t = y_{t-1} + \Delta y_t = 34 + 0.5818 = 34.5818$$

4.2.4 Evaluasi Model ARIMA

Evaluasi diterapkan dalam mengukur tingkat keakuratan model dalam prediksi nilai PM2.5. Berikut adalah metrik evaluasi yang diterapkan:

Tabel 8. Evaluasi Model ARIMA

Model	MSE	RMSE	MAPE (%)	MAE	N-BIC	Ljung-Box Q-Test
ARIMA (0,1,2)	4.318	2.078	25.004	1.656	1.595	0.542

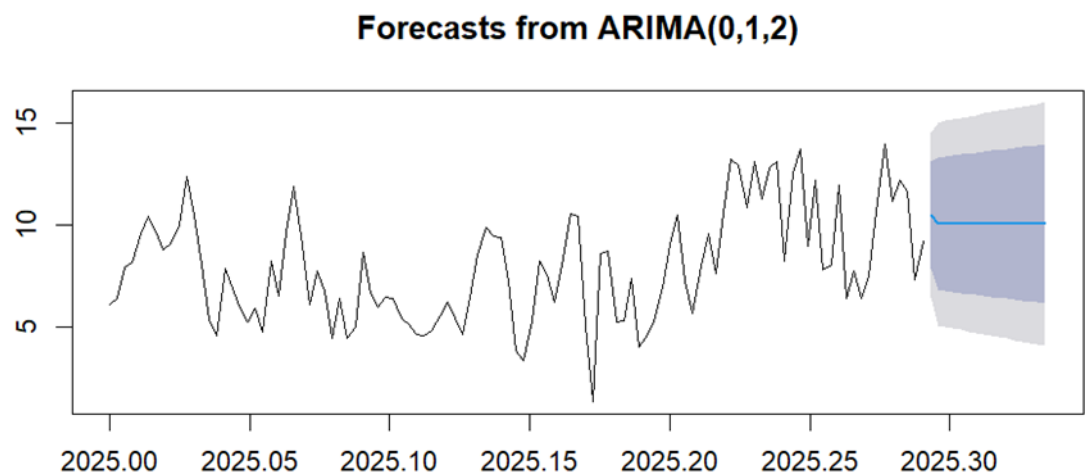
		Estimate	SE	t	Sig.
Constant		.022	.045	.496	.621
Difference		1			
MA	Lag 1	.290	.086	3.364	.001
	Lag 2	.501	.087	5.793	.000

Fig 5. Signifikasi Model ARIMA

4.2.5 Visualisasi ARIMA

Visualisasi diterapkan untuk memerhatikan dan melihat tren pola historis dan prediksi hasil model. Plot yang ditampilkan mencakup:

1. Grafik *time series* asli
2. Plot hasil peramalan 16 hari ke depan
3. Interval kepercayaan (*confidence interval*)



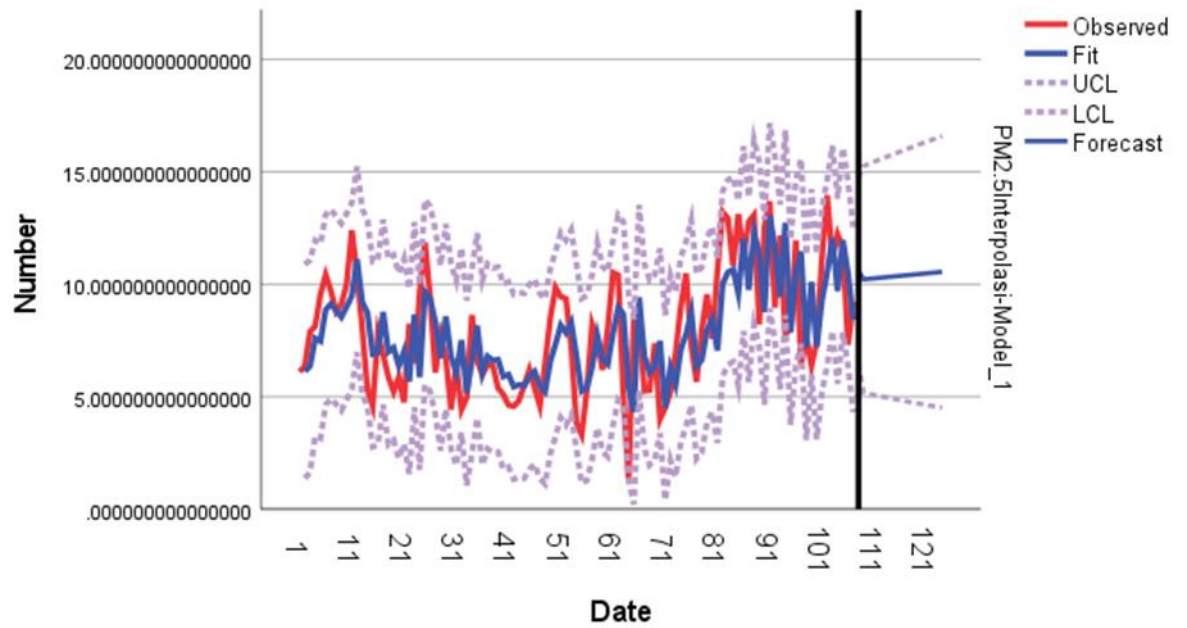


Fig 6. Visualisasi ARIMA

Hasil visualisasi yang diperoleh menunjukkan bahwa model mampu untuk mengikuti pola data historis dengan akurat, dan prediksi berada dalam batas kepercayaan yang wajar.

4.3 Implementasi ARIMAX

4.3.1 Struktur Model ARIMAX

Pembuatan model ARIMAX dirancang di R Studio, hal ini dilakukan untuk mempercepat waktu pengerjaan, mengurangi kesalahan, dan menyederhanakan proses analisis. Berikut adalah flowchart dari program.

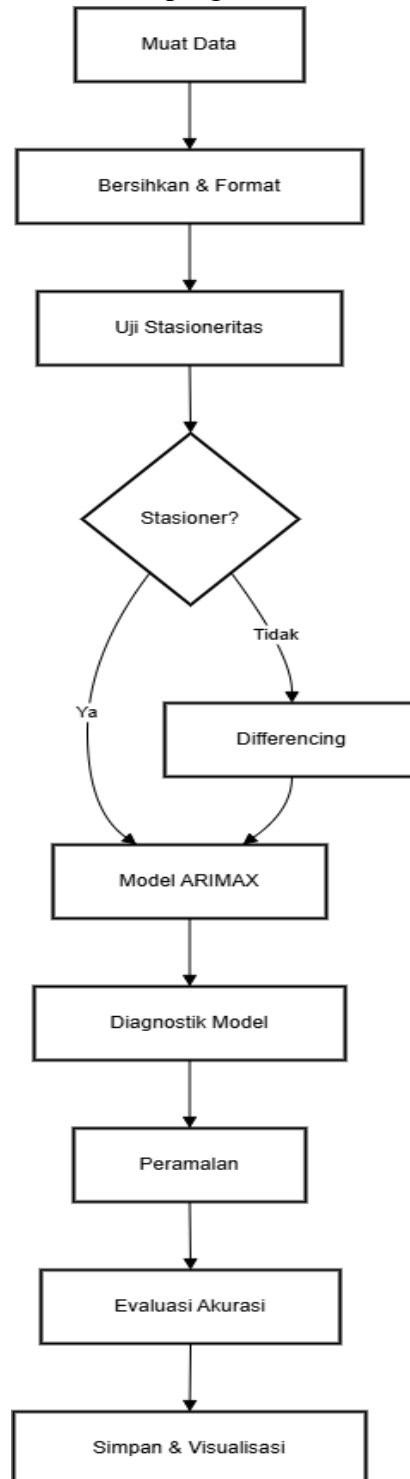


Fig 7. Flowchart ARIMAX

Proses diawali dengan memuat library yang diperlukan, kemudian mempersiapkan data. Jika data sudah bersih dilakukan uji stasioneritas menggunakan KPSS. Jika data PM2.5 belum stasioner, lakukan differencing hingga stasioner. Selanjutnya, dibangun model ARIMAX dengan variabel prediktor seperti suhu rata-rata (TAVG), kelembapan (RH_AVG), dan kecepatan angin (FF_AVG). Hasil model ditampilkan dalam ringkasan, lalu residualnya diuji menggunakan plot dan uji Ljung-Box untuk memastikan tidak ada masalah.

Untuk evaluasi kinerja model, data dibagi menjadi 80% data latih dan 20% data uji. Model dilatih pada data latih, kemudian prediksi dilakukan pada data uji. Akurasi model dihitung dan dibandingkan antara nilai aktual dan prediksi. Terakhir, hasil analisis disimpan ke dalam file Excel untuk dokumentasi.

4.3.2 Pembangunan Model ARIMAX

1. Uji Stasioneritas dan Differencing

Setelah data dibersihkan dan sebelum membangun model ARIMAX, penting untuk memastikan bahwa data yang digunakan telah bersifat stasioner, yaitu memiliki rata-rata dan varians yang konstan sepanjang waktu. Oleh karena itu, dilakukan uji Augmented Dickey-Fuller (ADF) terhadap variabel target PM2.5.

Dalam konteks uji ADF (Augmented Dickey-Fuller):

- Hipotesis nol (H_0): Data memiliki *unit root*, artinya tidak stasioner.
- Hipotesis alternatif (H_1): Data tidak memiliki unit root, artinya stasioner.

Kriteria pengambilan keputusan:

- Jika nilai test-statistic lebih kecil daripada critical value pada tingkat signifikansi yang ditentukan ($5\% = -3.43$), maka hipotesis nol (H_0) ditolak → data dianggap stasioner.

Hasil ADF Test:

```
Value of test-statistic is: -4.732 7

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
```

Fig 8. Hasil Test ADF

Pada hasil pengujian:

- Test-statistic = -4.732
- Critical value pada tingkat signifikansi $5\% = -3.43$

Karena $-4.732 < -3.43$, maka:

- Hipotesis nol (H_0) ditolak.
- Yang berarti data tidak memiliki unit root → data PM2.5 sudah stasioner.
- Oleh karena itu, proses differencing tidak diperlukan.

2. Pemodelan ARIMAX

Setelah dipastikan bahwa data PM2.5 telah stasioner, maka proses pemodelan dapat dilanjutkan tanpa perlu melakukan differencing.

Selanjutnya, pembuatan model ARIMAX (*Autoregressive Integrated Moving Average with Exogenous Variables*), yaitu pengembangan dari model ARIMA yang mengikutsertakan variabel eksternal sebagai prediktor.

Variabel prediktor (eksogen) yang digunakan dalam model ARIMAX ini adalah:

- **TAVG** → Temperatur rata-rata
- **RH_AVG** → Kelembapan rata-rata
- **FF_AVG** → Kecepatan angin rata-rata

Penentuan parameter p,d,q ARIMAX yang dibangun akan menggunakan fungsi ARIMA() dari paket **fable**, dengan code sebagai berikut:

```
model <- data_ts %>%
  model(arima = ARIMA(PM25 ~ TAVG + RH_AVG + FF_AVG))
```

Fig 9. Fungsi ARIMAX

Output ARIMA():

```
              arima
              <model>
1 <LM w/ ARIMA(1,0,0) errors>
```

Fig 10. Output ARIMA

Oleh karena itu pada model kali ini akan menggunakan parameter p (*autoregressive*), d (*differencing*), q (*moving average*) yang sama dengan output fungsi.

Selain itu, model ARIMAX juga menghasilkan koefisien untuk masing-masing variabel sebagai berikut:

```

Series: PM25
Model: LM w/ ARIMA(1,0,0) errors

Coefficients:
      ar1      TAVG      RH_AVG
      0.5844   0.4426  -0.0397
s.e.      0.0796   0.1167   0.0378
      FF_AVG
      -0.4153
s.e.      0.3223

```

Fig 11. Model ARIMAX

Interpretasi:

- AR(1): PM2.5 hari ini dipengaruhi oleh nilai PM2.5 hari sebelumnya.
- Nilai koefisien AR(1)(0.584) menunjukkan adanya *persistensi* pada PM2.5, artinya polusi udara cenderung dipengaruhi oleh keadaan sebelumnya.
- TAVG (0.443) memiliki pengaruh positif → semakin panas udara, cenderung meningkatkan kadar PM2.5.
- RH_AVG(-0.0397) dan FF_AVG(-0.415) memiliki pengaruh negatif → kelembapan dan kecepatan angin membantu mengurangi konsentrasi PM2.5.

3. Diagnostik Residual

Setelah model ARIMAX dibangun, langkah selanjutnya adalah diagnostik residual untuk memastikan:

- Tidak ada pola sistematis dalam residual,
- Residual bersifat acak (*white noise*), dan
- Tidak ada autokorelasi signifikan.

Langkah diagnostik:

- 1) Visualisasi diagnostik residual
- 2) Uji Ljung-Box → untuk menguji autokorelasi residual secara statistik.

a. Output Diagnostik Residual

1) Visualisasi diagnostik residual

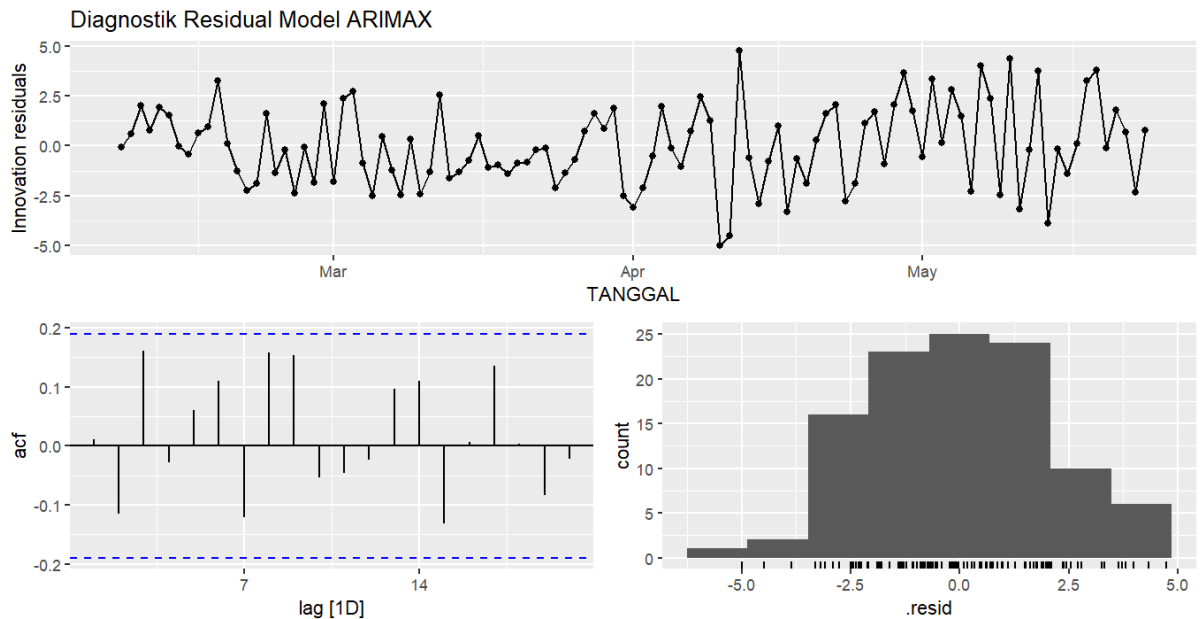


Fig 12. Visualisasi diagnostik residual

- **Plot Residual vs. Waktu**
Menunjukkan bahwa residual model ARIMAX berfluktuasi secara acak di sekitar nilai nol (-5.0 hingga 5.0) tanpa pola tren atau musiman yang jelas. Meskipun terdapat beberapa outlier (sekitar bulan April), secara keseluruhan tidak ada pola sistematis yang tertinggal, menandakan bahwa model telah menangkap mayoritas struktur data dengan baik.
- **Plot ACF (Autocorrelation Function)**
Mengonfirmasi tidak adanya autokorelasi signifikan pada residual, dengan nilai lag (1D) yang seluruhnya berada dalam batas signifikansi (-0.2 hingga 0.2). Hal ini memenuhi kriteria residual sebagai *white noise*, yang berarti tidak ada ketergantungan waktu yang belum termodelkan.
- **Histogram Residual**
Memperlihatkan distribusi yang simetris dengan mean mendekati nol. Tidak ada indikasi deviasi ekstrem dari normalitas, sehingga model dianggap memadai. Secara keseluruhan, diagnostik residual mengindikasikan bahwa model ARIMAX ini valid dan siap digunakan untuk analisis lebih lanjut.

b. Uji Ljung-Box

```
Hasil Uji Ljung-Box:
> print(ljung_box_test)
# A tibble: 1 × 3
  .model lb_stat lb_pvalue
  <chr>   <dbl>   <dbl>
1 arima    14.1    0.169
```

Fig 13. Hasil Uji Ljung-Box

Nilai statistik uji (14.1) terlihat cukup tinggi, namun p value (0.169) yang lebih besar dari tingkat signifikansi 0.05 menunjukkan tidak adanya autokorelasi yang signifikan dalam residual. Hal ini berarti kita tidak memiliki bukti statistik yang cukup untuk menolak hipotesis null bahwa residual bersifat acak (*White Noise*). Dengan demikian, model ini dapat dianggap valid dan tidak memerlukan revisi lebih lanjut terkait struktur autokorelasi.

4.3.3 Evaluasi Model ARIMAX

Setelah pembangunan model ARIMAX selesai, dilakukan proses evaluasi untuk menilai performa model dalam memprediksi kadar PM2.5 berdasarkan variabel cuaca yang berpengaruh, yaitu TAVG (temperatur rata-rata), RH_AVG (kelembaban rata-rata), dan FF_AVG (kecepatan angin rata-rata).

Hasil Analisis Model:

	.model	.type	ME	RMSE	MAE	MPE	MAPE
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	ARIMA(P...	Test	2.45	3.46	2.76	19.1	23.8

Fig 14. Hasil Analisis Model

- MAE (Mean Absolute Error) = 2.76
- RMSE (Root Mean Squared Error) = 3.46
- MAPE (Mean Absolute Percentage Error) = 23.8%

Berdasarkan hasil evaluasi model ARIMAX terhadap data PM2.5, diperoleh nilai RMSE sebesar 3.46, MAE sebesar 2.76, dan MAPE sebesar 23.8%. Nilai RMSE menunjukkan bahwa secara rata-rata, kesalahan prediksi model berada sekitar ± 3.46 dari nilai aktual. Sementara itu, nilai MAE yang sebesar 2.76 memperlihatkan bahwa kesalahan absolut rata-rata model juga cukup rendah dan konsisten dengan RMSE, yang menandakan tidak adanya kesalahan ekstrem dalam prediksi. Selain itu, nilai MAPE sebesar 23.8% menunjukkan bahwa rata-rata kesalahan model terhadap nilai aktual adalah sekitar 24% masih tergolong wajar. Secara keseluruhan, hasil ini menunjukkan bahwa model ARIMAX memiliki performa yang cukup baik dan dapat diandalkan untuk memprediksi konsentrasi PM2.5.

4.3.4 Visualisasi ARIMAX

Berikut adalah visualisasi perbandingan Nilai aktual vs Nilai Prediksi pada model ARIMAX.

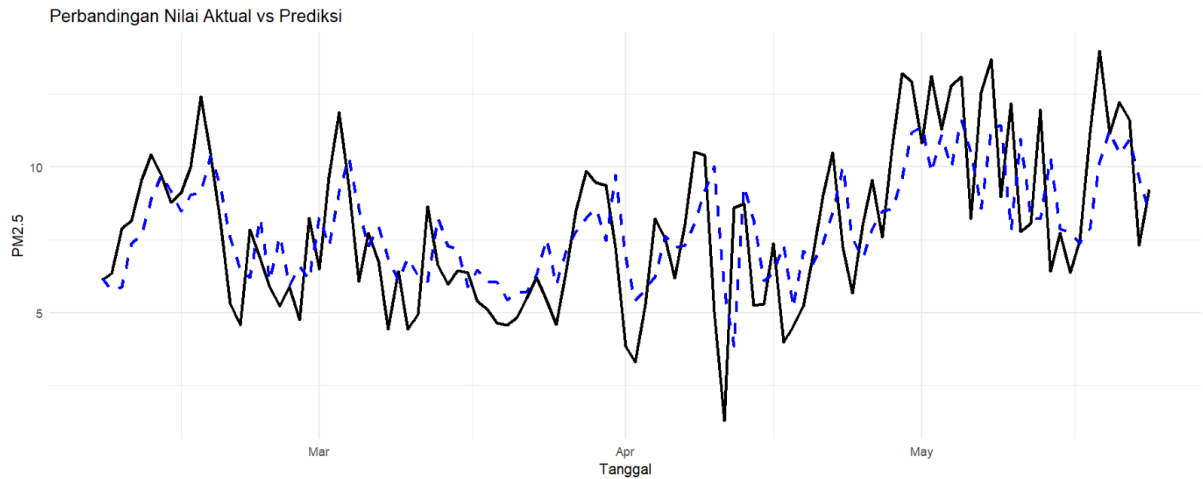


Fig 15. Visualisasi ARIMAX

Hasil prediksi dibandingkan dengan data aktual menggunakan grafik garis. Garis hitam mewakili nilai aktual PM2.5, sedangkan garis biru putus-putus menunjukkan nilai yang diprediksi oleh model. Hasil visual menunjukkan bahwa model mampu menangkap pola naik-turun PM2.5 dengan cukup baik, meskipun terdapat sedikit penyimpangan pada beberapa titik ekstrem.

4.4 ARIMA vs ARIMAX

Setelah permodelan menggunakan pendekatan ARIMA, yaitu univariat dan ARIMAX, yaitu multivariat dengan variabel eksogen. Selanjutnya, melakukan perbandingan kinerja antara kedua model dalam peramalan konsentrasi PM2.5. Perbandingan yang dilakukan bertujuan untuk tahu apakah penambahan variabel eksternal memberikan peningkatan performa dibanding dengan model ARIMA murni yang hanya menggunakan data PM2.5 historis. Untuk melakukan evaluasi, diterapkan perbandingan model berdasarkan MSE, RMSE, dan MAPE. Berikut adalah tabel untuk perbandingan kinerja model:

Tabel 9. Evaluasi Model ARIMA dan ARIMAX

Model	MSE	RMSE	MAPE (%)	Catatan Evaluasi
ARIMA (0,1,2)	4.318	2.078	25.004	Model univariat, residual white noise, & MA signifikan.
ARIMAX (1,0,0)	11.97	3.46	23.8	Model multivariat dan mempertimbangkan faktor cuaca

BAB V

KESIMPULAN

5.1 KESIMPULAN

Hasil penelitian dan analisis terhadap konsentrasi PM2.5 menggunakan model ARIMA dan ARIMAX sehingga dapat disimpulkan berikut:

- Model ARIMA (0,1,2) berhasil dibentuk setelah melalui proses pengujian stasioneritas dan proses pola identifikasi ACF dan PACF. Model telah memberikan performa prediksi yang cukup baik dengan nilai RMSE sebesar 2.078, MAPE sebesar 25.004 %, dan N-BIC sebesar 1.595, serta menghasilkan residual yang bersifat *white-noise* berdasarkan uji Ljung-Box.
- Pemodelan ARIMAX (1,0,0) telah dikembangkan dengan menambahkan variabel eksogen. Hasil yang diperoleh menunjukkan bahwa model ARIMAX memiliki performa yang sedikit lebih baik dengan RMSE 3.46 dan MAPE 23.8% yang berarti bahwa faktor cuaca berpengaruh terhadap fluktuasi PM2.5.
- Secara menyeluruh, ARIMAX lebih unggul daripada ARIMA meski selisih performa yang tidak terlalu signifikan. Namun secara konsep, ARIMAX representatif untuk data lingkungan yang dipengaruhi oleh berbagai faktor.
- Metode hybrid mean-interpolation dalam menangani *missing value* cukup efektif.
- Hasil dari visualisasi menunjukkan bahwa model mampu mengikuti tren historis dengan cukup baik dan prediksi ke depan berada dalam batas yang dapat diterima.

5.2 SARAN

Berdasarkan hasil penelitian dan implementasi dari model ARIMA dan ARIMAX untuk melakukan peramalan konsentrasi PM2.5, terdapat beberapa saran untuk pengembangan lebih lanjut:

- Dalam meningkatkan akurasi prediksi model ARIMAX, disarankan menambahkan variabel eksternal lainnya yang diperoleh.
- Disarankan menggunakan metode validasi silang berbasis waktu.
- Model prediksi yang lebih kompleks dapat diterapkan untuk melihat potensi peningkatan akurasi.
- Disarankan menggunakan sistem otomatisasi deteksi dan imputasi nilai hilang.
- Integrasi model ke dalam sistem monitoring kualitas udara secara *real-time*.
- Hasil prediksi dapat dikembangkan dalam bentuk yang lebih interaktif.

DAFTAR PUSTAKA

- Akbar, M. (2016). *BAB II Tinjauan Pustaka*. Repository ITK.
- Basuki, A., & Prawoto, H. (2016). *BAB III Metode Penelitian*. Repository Untidar.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day.
- Digilib Unila. (2017). *Uji kestasioneran mengenai data inflasi*. Universitas Lampung.
- Hyndman, R. J. (2022). Discussion on ARIMAX model and exogenous variables in air quality forecasting.
- Jusri. (2023). *Makassar city air pollution forecast in 2045*. Hasanuddin University.
- Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi). (2024). Forecasting the air quality index by utilizing several meteorological factors using ARIMAX method.
- Keramidas, T., Tchung-Ming, S., Weitzel, M., & Van Dingenen, R. (2023). Analisis kualitas udara dan dampak PM2.5 terhadap kesehatan masyarakat. *Jurnal Ilmu Lingkungan*, 21(2), 381–386.
- Kuncoro, M. (2017). *Metode riset untuk bisnis dan ekonomi*. Erlangga.
- Moshhammer, H., & Wallner, P. (2011). Air quality as respiratory health indicator—A critical review. *International Journal of Occupational Medicine and Environmental Health*, 24(3), 241–248. <https://doi.org/10.2478/s13382-011-0028-9>
- Pope, C. A., III, & Dockery, D. W. (2006). Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association*, 56(6), 709–742. <https://doi.org/10.1080/10473289.2006.10464485>
- Ramdhan, A., & Syamsuddin, S. (2020). Urban growth and industrial development in Makassar City: Impacts on environmental sustainability. *Journal of Environmental Management and Tourism*, 11(4), 789–796.
- Repository UIN Suska. (2022). *BAB III*. Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Utama, D. A. (2019). Indeks standar pencemar udara polutan karbon monoksida di Terminal Malengkeri Kota Makassar. *Jurnal UNHAS*.
- Wahid, S., & Setyawan, H. (2020). Peramalan data deret waktu menggunakan model ARIMA. *Indonesian Journal of Applied Statistics*, 3(2), 136–147. <https://jurnal.uns.ac.id/ijas/article/view/26298/19416>
- Widarjono, A. (2018). *Ekonometrika: Teori dan aplikasi untuk ekonomi dan bisnis* (Edisi ke-3). UPP STIM YKPN.
- World Health Organization. (2021). *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World

Health Organization.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting* (2nd ed.). John Wiley & Sons.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>